AD-A046 250 MARYLAND UNIV COLLEGE PARK DE'T OF MATHEMATICS F/G 12/1
ESTIMATION OF THE PARAMETERS IN STATIONARY AUTOREGRESSIVE PROCE--ETC(U)
AUG 77 B KEDEM AFOSR-77-3332
UNCLASSIFIED MD77-54-BK AFOSR-TR-77-1231 NL

| OF |
AD
A046250

END
DATE
FILMED
12-77
DDC

(18) AFOSR (19) TR-77-1231

(5)

AD A046250

(6) ESTIMATION OF THE PARAMETERS IN STATIONARY AUTOREGRESSIVE PROCESSES AFTER HARD LIMITING.

by

(10) Benjamin Kedem

(9) Interim rept.,

(15) AFOSR-77-3332

DDC
RECEIVED
NOV 2 1977
F

Department of Mathematics
University of Maryland
College Park, Maryland
20742

(16) 2304 (17) A5

AD No.
DDC FILE COPY

Approved for public release; dist...

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. SPECIAL
A

(14) MD77-54-BK, TR77-42

(11) Aug[redacted]77 (12) 25p.

400 953

LB

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

ESTIMATION OF THE PARAMETERS IN STATIONARY AUTOREGRESSIVE PROCESSES AFTER HARD LIMITING

by

Benjamin Kedem

§1. Introduction

Let Z_t, $t = 0, \pm 1, \ldots,$ be a stationary stochastic process and define the clipped process X_t by

$$X_t = \begin{cases} 1, & Z_t \geq 0 \\ 0, & Z_t < 0 \end{cases}, \qquad t = 0, \pm 1, \ldots \tag{1.1}$$

There are several reasons for the transformation (1.1) commonly known as clipping or hard limiting. If long records of data are available then it is convenient to clip the original data and store the observed information in just a few sufficient [illegible]tics whose computation is extremely fast as only counting is in[illegible]ved. All the more so, experience has shown, [6], [7], [10], [11], [13], [18] that very little precision in estimation is lost due to this coarse quantization. Also, there are situations [9] when the data must be actually observed in the form (1.1) as to allow certain noise modulation. Evidently the problem of outliers in the Z_t data ceases to exist.

From a statistical viewpoint (1.1) is tractable because: if the binary process X_t is finitely dependent then the likelihood function of an observed time series $X_1, \ldots, X_n$ is always available regardless of the probability law which governs the original process Z_t. This means that efficient estimation of parameters based on the quantized data is possible.

Suppose now Z_t is an autoregressive process of order p

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + u_t, \quad t = 0, \pm 1, \ldots, \tag{1.2}$$

where the u_t are independent $N(0,\sigma^2)$ variates, and where the roots of the associated polynomial $1 - \phi_1 B - \ldots - \phi_p B^p$ lie outside the unit circle. In this paper we shall obtain approximate maximum likelihood estimates for $\phi_1, \ldots, \phi_p$ from the clipped process X_t, when X_t is a stationary p^{th}-order Markov chain. The related problems of estimating the covariance function and spectral density of Z_t from the clipped data were discussed in [6], [11], [13], [17], [18]. The main difference between these works and the present one is that we base our estimation on the likelihood function of the 0-1 data where the sufficient statistics are determined by the <u>axis-crossings</u> by Z_t. In other words the estimates are functions of the particular arrangements and lengths of the 1-runs and 0-runs as formed by Z_t while crossing level 0. This principle will become clear as we proceed. It is shown that the clipped data estimates are obtained at a substantial saving in computing as expressed in terms of the number of arithmetical operations. At the same time they compete well with the usual maximum likelihood estimates.

Experience has shown that (1.2) is an adequate representation for describing a wide range of stationary time series e.g. see [3], [14]. Maximum likelihood estimation of the parameters in (1.2) based on the original data Z_t has been discussed in [12]. Moderate modifications are considered in [1], [3]. It turns out that the maximum likelihood estimates of the ϕ_j ob-

obtained from the original data are essentially the same as the least squares estimates.

In what follows we shall make use of a famous formula. Let (X,Y) have a bivariate normal distribution with parameters $EX = EY = 0$, $\operatorname{Var} X = \operatorname{Var} Y = \sigma^2$ and correlation ρ. Then by introducing polar coordinates one obtains

$$\Pr(X \geq 0,\ Y \geq 0) = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}(\rho). \tag{1.3}$$

This formula can be traced back to the writings of Sheppard [15] Stieltjes [16].

§2. The Likelihood Function of the Clipped Time Series and Estimation of Transition Probabilities

If X is a Bernoulli random variable with a probability of success θ then $\Pr(X = x) = \theta^x(1 - \theta)^{1-x}$, $x = 0,1$. In deriving this exponential distribution, use was made of the fact that X can take the values 0 or 1. This observation is instrumental in constructing the likelihood function of the clipped series X_t.

Define

$$\Pr(X_i = x_i \mid X_{i-1} = x_{i-1}, \ldots, X_{i-k} = x_{i-k}) = P_{x_i x_{i-1} \cdots x_{i-k}}$$

and observe that from (1.1) we have

$$\Pr(X_i = x_i) = \frac{1}{2},$$

$$\Pr(X_i = x_i \mid X_{i-1} = x_{i-1}) = \left[p_{11}^{x_i}(1-p_{11})^{1-x_i}\right]^{x_{i-1}} \cdot \left[p_{10}^{x_i}(1-p_{10})^{1-x_i}\right]^{1-x_{i-1}},$$

$$\Pr(X_i = x_i \mid X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}) = \left\{\left[p_{111}^{x_i} p_{011}^{1-x_i}\right]^{x_{i-1}} \cdot \left[p_{101}^{x_i} p_{001}^{1-x_i}\right]^{1-x_{i-1}}\right\}^{x_{i-2}}$$
$$\left\{\left[p_{110}^{x_i} p_{010}^{1-x_i}\right]^{x_{i-1}} \cdot \left[p_{100}^{x_i} p_{000}^{1-x_i}\right]^{1-x_{i-1}}\right\}^{1-x_{i-2}},$$

etc., and therefore in general

$$\Pr(X_1 = x_1,\ldots,X_n = x_n) = \frac{1}{2} \prod_{i=2}^{n} \prod p_{y_i y_{i-1}\cdots y_1}^{I_i I_{i-1}\cdots I_1}, \qquad (2.1)$$

such that the second product is over all 2^i 0-1 i-tuples $(y_i, y_{i-1},\ldots,y_1)$ and

$$I_i = \begin{cases} x_i & \text{if } y_i = 1 \\ 1-x_i & \text{if } y_i = 0. \end{cases}$$

This is the joint distribution of any 0-1 time series $X_1,\ldots,X_n$ provided $\Pr(X_1 = x_1) = \frac{1}{2}$.

Evidently, since Z_t, the original process, is strictly stationary also the clipped process is strictly stationary; that is, the stationary property was preserved by X_t. But Z_t is also Markovian. However, mathematically speaking, X_t is not a Markov process as well because for one thing, the hard limiting transformation is not 1-1. In this case it is apparent that (2.1) viewed as a likelihood function is useless, for as the series size n increases also the number of unknown parameters increases so that (2.1) cannot be used for estimation purposes. To get around this difficulty one has to condense somehow (2.1). We shall then overlook the mathematical difficulty and assume that the pth order Markov property is inherent in the clipped process. This is a statistical assumption which is certainly reasonable when σ^2 is small. This assumption is analogous to the normal assumption in the general linear model or to the

assumption of Poisson arrivals in queuing theory for it enables us to pursue statistical analysis based on the likelihood (2.1).

Now let Z_t, $t = 1,2,\ldots,n$, be given by (1.2), and X_t, $t = 1,2,\ldots,n$, be given by (1.1). Under the assumption that X_t is a pth order Markov process the joint distribution (2.1) reduces to

$$\Pr(X_1 = x_1,\ldots,X_n = x_n) \propto \prod_{i=p+1}^{n} \prod p_{y_i y_{i-1}\cdots y_{i-p}}^{I_i I_{i-1}\cdots I_{i-p}} \tag{2.2}$$

$$= \prod p_{y_i y_{i-1}\cdots y_{i-p}}^{\sum_{i=p+1}^{n} I_i I_{i-1}\cdots I_{i-p}},$$

where as before the product in the last expression extends over all the 2^{p+1} 0-1 arrangements $(y_i, y_{i-1},\ldots,y_{i-p})$. In (2.2) terms which do not depend on n were ignored. This likelihood function is a special case of the one given by Bartlett (1951), except that formally we utilize it in a different way. We shall make use of the following identity:

$$\begin{aligned} 1 &= (1-z) + z = (1-y)(1-z) + y(1-z) + (1-y)z + yz \\ &= (1-x)(1-y)(1-z) + x(1-y)(1-z) + (1-x)y(1-z) + xy(1-z) \\ &\quad + (1-x)(1-y)z + x(1-y)z + (1-x)yz + xyz, \end{aligned} \tag{2.3}$$

and so on. Define

$$\lambda_k = \Pr(X_t = 1 \mid X_{t-k} = 1), \quad k = 1,2,\ldots,p.$$

We estimate the ϕ_k from the λ_k. It is possible to rewrite (2.2) as a product of a function in λ_k and a function of nuisance parameters in this case higher order probabilities. The function in λ_k is a power of λ_k times a power of $1 - \lambda_k$. Only terms in (2.2) of the form

$$\sum_{i=p+1}^{n} I_i \cdots I_{i-p+k+1} x_{i-p+k} I_{i-p+k-1} \cdots I_{i-p+1} x_{i-p} \qquad (2.4)$$
$$p_{y_i \cdots y_{i-p+k+1}\, 1\, y_{i-p+k-1} \cdots y_{i-p+1}\, 1}$$

or of the same form but with the x's replaced by $(1-x_{i-p+k})$ and $(1-x_{i-p})$ respectively, and the 1's replaced by 0's, contribute their exponents to the power of λ_k. The rest of the terms in (2.2) contribute their exponents to $(1 - \lambda_k)$. The sum of all these exponents in n-p. Thus the exponent of λ_k is

$$\sum_{i=p+1}^{n} [\sum I_i \cdots I_{i-p+k+1} I_{i-p+k-1} \cdots I_{i-p+1}] x_{i-p+k} x_{i-p}$$

$$+ \sum_{i=p+1}^{n} [\sum I_i \cdots I_{i-p+k+1} I_{i-p+k-1} \cdots I_{i-p+1}](1-x_{i-p+k})(1-x_{i-p}),$$

where the second sum extends over all possible terms of the form (2.4). Evidently, this second sum is just the telescopic identity (2.3) and is equal to 1. The exponent of $1 - \lambda_k$ is obtained in the same way. (2.2) can now be rewritten as

$$\Pr(X_1 = x_1,\ldots,X_n = x_n) \propto \lambda_k^{2\sum_{i=p+1}^{n} x_{i-p+k}x_{i-p} - \sum_{i=p+1}^{n}(x_{i-p+k}+x_{i-p})+(n-p)} \tag{2.5}$$

$$\cdot(1-\lambda_k)^{-2\sum_{i=p+1} x_{i-p+k}x_{i-p} + \sum_{i=p+1}^{n}(x_{i-p+k}+x_{i-p})}$$

$$\cdot \text{ powers of nuisance parameters.}$$

It follows that a maximum likelihood estimate of λ_k is given by

$$\hat{\lambda}_k = \frac{\sum_{i=p+1}^{n} x_{i-p+k}x_{i-p} - \sum_{i=p+1}^{n}(x_{i-p+k}+x_{i-p}) + (n-p)}{n-p}, \quad k = 1,2,\ldots,p. \tag{2.6}$$

We h[illegible] $E\hat{\lambda}_k = \lambda_k$ and for a sufficiently large n [2] $\hat{\lambda}_k$ is asymptotically normal such that $\mathrm{Var}\,\sqrt{n}(\hat{\lambda}_k-\lambda_k) \sim \lambda_k(1-\lambda_k)$.

From a computational viewpoint it is preferable to express $\hat{\lambda}_k$ in the simpler form

$$\hat{\lambda}_k = \frac{2R_k - 2S + n}{n-k}, \quad k = 1,2,\ldots,p, \tag{2.7}$$

where $R_k = \sum_{i=k+1}^{n} X_iX_{i-k}$, $S = \sum_{i=1}^{n} X_i$. This unbaised form is essentially the same as (2.6) except for end effects of the series which become negligible as n increases. We shall use (2.7). In the next section we show how to obtain approximate maximum likelihood estimates for the ϕ_k from the λ_k.

<u>Now we immediately see that</u> $\hat{\lambda}_k$ <u>depends on the axis crossings by</u> Z_t. In particular

$$\hat{\lambda}_1 = \frac{n-2(\#\text{ of 1-runs})}{n-1}, \tag{2.8}$$

or, neglecting end effects as is the case for large n

$$\hat{\lambda}_1 = \frac{n - (\#\text{ of axis crossings by } Z_t)}{n-1} \tag{2.9}$$

$$\hat{\lambda}_2 = \frac{n-2(\#\text{ of runs between the first and last 1 with two or more symbols} + 1)}{n-2} \tag{2.10}$$

etc.

The exact distribution of the number of axis crossings by Z_t when X_t is either a first or second order Markov chain has been recently obtained in [8].

It should be noted that the estimation problem of transition probabilities treated here is somewhat different than the usual problem of estimation of conditional probabilities in p^{th} order Markov chains where the quantities of interest are the probabilities of a state given the previous p states. That problem was first solved in [2].

§3. Estimation of the ϕ's

We start with a brief review of the maximum likelihood estimation of $\phi_1,\ldots,\phi_p$ from the original Z series. This will give rise to natural estimates based on the clipped data.

Given a time series $z_1,\ldots,z_n$ from the Gaussian process Z_t then its joint distribution is $N(\underset{\sim}{0},\underset{\sim}{\Gamma}_n)$, where $\underset{\sim}{\Gamma}_n = \{\gamma_{|i-j|}\}$, $i,j = 1,\ldots,n$, and the joint density can be written as the product

$$f(z_1,\ldots,z_n|\underset{\sim}{\phi},\sigma^2) = f(z_{p+1},\ldots,z_n|\underset{\sim}{\phi},\sigma^2,z_1,\ldots,z_p)f(z_1,\ldots,z_p|\underset{\sim}{\phi},\sigma^2).$$

Because the second term on the right hand side is independent of n the log-likelihood for sufficiently large n can be conveniently defined as

$$\begin{aligned}\ell(\underset{\sim}{\phi}) &= \log f(z_{p+1},\ldots,z_n|\underset{\sim}{\phi},\sigma^2,z_1,\ldots,z_p)\\ &= -\frac{n-p}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(1,\underset{\sim}{\phi}')\begin{bmatrix}\Sigma z_t^2 & -\Sigma \underset{\sim}{z}_t' z_t\\ -\Sigma \underset{\sim}{z}_t z_t & \Sigma \underset{\sim}{z}_t \underset{\sim}{z}_t'\end{bmatrix}\begin{pmatrix}1\\ \underset{\sim}{\phi}\end{pmatrix} \qquad (3.1)\end{aligned}$$

where $\underset{\sim}{\phi} = (\phi_1,\ldots,\phi_p)'$, $\underset{\sim}{z}_t = (z_{t-1},z_{t-2},\ldots,z_{t-p})'$, and the sums extend from $p+1$ to n. Then

$$\frac{\partial}{\partial\underset{\sim}{\phi}}\ell(\underset{\sim}{\phi}) = \frac{1}{2\sigma^2}\{2\Sigma\,\underset{\sim}{z}_t z_t - 2\Sigma\,\underset{\sim}{z}_t\underset{\sim}{z}_t'\underset{\sim}{\phi}\}$$

and by equating to $\underset{\sim}{0}$ we obtain the maximum likelihood estimator

$$\tilde{\underset{\sim}{\phi}} = (\Sigma \underset{\sim}{z}_t \underset{\sim}{z}_t')^{-1}\,\Sigma \underset{\sim}{z}_t z_t \qquad (3.2)$$

Now, multiply (3.2) by $1 = \Sigma z_t^2/\Sigma z_t^2$ and <u>define</u>

$$\tilde{\lambda}_k = \frac{1}{2} + \frac{1}{\pi}\sin^{-1}\left(\frac{\Sigma z_t z_{t-k}}{\Sigma z_t^2}\right).$$

Then apart from neglibible end effects (3.2) takes the form

$$\tilde{\underset{\sim}{\phi}} = \underset{\sim}{\phi}(\tilde{\underset{\sim}{\lambda}}) = \begin{bmatrix}1 & \sin\pi(\tilde{\lambda}_1-\tfrac{1}{2}) & \ldots & \sin\pi(\tilde{\lambda}_{p-1}-\tfrac{1}{2})\\ \sin\pi(\tilde{\lambda}_1-\tfrac{1}{2}) & 1 & \ldots & \sin\pi(\tilde{\lambda}_{p-2}-\tfrac{1}{2})\\ \vdots & \vdots & & \vdots\\ \sin\pi(\tilde{\lambda}_{p-1}-\tfrac{1}{2}) & \sin\pi(\tilde{\lambda}_{p-2}-\tfrac{1}{2}) & \ldots & 1\end{bmatrix}^{-1}\begin{bmatrix}\sin\pi(\tilde{\lambda}_1-\tfrac{1}{2})\\ \sin\pi(\tilde{\lambda}_2-\tfrac{1}{2})\\ \vdots\\ \sin\pi(\tilde{\lambda}_p-\tfrac{1}{2})\end{bmatrix}, \qquad (3.3)$$

where $\tilde{\underset{\sim}{\lambda}} = (\tilde{\lambda}_1,\ldots,\tilde{\lambda}_p)'$. It follows that the approximate maximum likelihood estimator based on the clipped data is

$$\hat{\underset{\sim}{\phi}} = \underset{\sim}{\phi}(\hat{\underset{\sim}{\lambda}}) \tag{3.4}$$

where $\hat{\lambda}_k$ is given in (2.7) and $\hat{\underset{\sim}{\lambda}} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_p)'$. It is possible to calculate $\underset{\sim}{\phi}(\hat{\underset{\sim}{\lambda}})$ recursively using formulae developed in [5]. This is equivalent to finding the inverse in (3.3).

From (3.3) and (3.4) it is seen that the estimation problem of $\underset{\sim}{\phi}$ was reduced to the estimation of $\underset{\sim}{\lambda}$ and that the loss of efficiency due to clipping depends on the closeness of $\hat{\underset{\sim}{\lambda}}$ to $\tilde{\underset{\sim}{\lambda}}$. Obviously when $\hat{\underset{\sim}{\lambda}}$ and $\tilde{\underset{\sim}{\lambda}}$ are close so are $\hat{\underset{\sim}{\phi}}$ and $\tilde{\underset{\sim}{\phi}}$ by continuity. It is convenient to introduce the shift operator B defined by $BZ_t = Z_{t-1}$, and let $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$. Also let ρ_k be the correlation function of Z_t and let $\kappa_x(i,j,k)$ be the fourth order cumulant function of X_t (obviously it corresponds to stationarity).

$$\begin{aligned}\kappa_x(i,j,k) = {} & E(X_t-\tfrac{1}{2})(X_{t+i}-\tfrac{1}{2})(X_{t+j}-\tfrac{1}{2})(X_{t+k}-\tfrac{1}{2}) \\ & - \frac{1}{4\pi^2}\{\sin^{-1}(\rho_i)\sin^{-1}(\rho_{j-k}) \\ & + \sin^{-1}(\rho_j)\sin^{-1}(\rho_{i-k}) + \sin^{-1}(\rho_k)\sin^{-1}(\rho_{i-j})\}.\end{aligned}$$

Because Z_t is Gaussian κ_Z vanishes. Now κ_x is a measure of dependence in the clipped series and since the pairwise dependence in this series is weaker than the pairwise dependence in the original Gaussian process (see (4.1) and the remarks following it) it is reasonable to assume that κ_X is small and summable.

<u>Theorem 1</u>: Let Z_t be an AR(p) process

$$\phi(B)Z_t = u_t, \quad t = 0, \pm 1, \ldots,$$

u_t are independent $N(0,\sigma^2)$ random variables, where the roots of $\phi(B) = 0$ lie outside the unit circle. Let X_t be given by (1.1). If $\left| \sum_{r=-\infty}^{\infty} \kappa_X(k,-r,j-r) \right| < \infty$ then $\hat{\underset{\sim}{\phi}}$ converges to $\underset{\sim}{\phi}$ in mean square in the sense that

$$\lim_{n \to \infty} E\|\hat{\underset{\sim}{\phi}} - \underset{\sim}{\phi}\| = 0,$$

where $\|\cdot\|$ denotes the usual norm.

<u>Proof</u>: With obvious notation write

$$\underset{\sim}{\phi} = \underset{\sim}{\phi}(\underset{\sim}{\lambda}) = \underset{\sim}{M}^{-1}\underset{\sim}{\rho}.$$

Let

$$\sin\pi(\hat{\lambda}_k - \tfrac{1}{2}) = \sin\pi(\lambda_k - \tfrac{1}{2}) + \varepsilon_k, \quad k = 1, \ldots, p,$$

$$\underset{\sim}{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)',$$

and

$$\underset{\sim}{E} = \begin{pmatrix} 0 & \varepsilon_1 & \varepsilon_2 \cdots \varepsilon_{p-1} \\ \varepsilon_1 & 0 & \varepsilon_1 \cdots \varepsilon_{p-2} \\ \varepsilon_2 & \varepsilon_1 & 0 \cdots \varepsilon_{p-2} \\ \vdots & \vdots & \vdots \quad \vdots \\ \varepsilon_{p-1} & \varepsilon_{p-2} & \varepsilon_{p-3} \cdots 0 \end{pmatrix}$$

Then

$$\underset{\sim}{\phi}(\hat{\underset{\sim}{\lambda}}) = (\underset{\sim}{M} + \underset{\sim}{E})^{-1}(\underset{\sim}{\rho}+\underset{\sim}{\varepsilon}).$$

Observe that $|\varepsilon_k| \le 2|\sin \frac{\pi}{2} (\hat{\lambda}_k-\lambda_k)|$. It follows from the fact that ρ_k is absolutely summable, $E\hat{\underset{\sim}{\lambda}} = \underset{\sim}{\lambda}$, and [1] p. 464,

$$\lim_{n\to\infty} n \operatorname{Cov}(\hat{\lambda}_k,\hat{\lambda}_j) = \frac{1}{\pi^2} \sum_{r=-\infty}^{\infty} [\sin^{-1}(\rho_r)\sin^{-1}(\rho_{r+k-j}) + \sin^{-1}(\rho_{r-j})\sin^{-1}(\rho_{r+k})]$$

$$+ 4 \sum_{r=-\infty}^{\infty} \kappa_x(k,-r,j-r) \qquad (3.5).$$

that $E\,\varepsilon_k^2 = O(1/n)$. Thus, for sufficiently large n we can use results from perturbation theory. Accordingly [4] we add another equation

$$\underset{\sim}{\phi}_0 = \underset{\sim}{M}^{-1}(\underset{\sim}{\rho}+\underset{\sim}{\varepsilon})$$

and define the norm of a $p \times p$ matrix $\underset{\sim}{A}$ by

$$\|\underset{\sim}{A}\| = \underset{\underset{\sim}{v} \in E^p}{\text{maximum}} \frac{\|\underset{\sim}{A}\underset{\sim}{v}\|}{\|\underset{\sim}{v}\|}$$

We shall find a bound on $\|\hat{\underset{\sim}{\phi}}-\underset{\sim}{\phi}\|$. Note that

$$\|\hat{\underset{\sim}{\phi}} - \underset{\sim}{\phi}\| \le \|\hat{\underset{\sim}{\phi}} - \underset{\sim}{\phi}_0\| + \|\underset{\sim}{\phi}_0 - \underset{\sim}{\phi}\|$$

As n increases $\|\underset{\sim}{M}^{-1}\underset{\sim}{E}\| = \delta = O_p(1/\sqrt{n})$ becomes small, so that for sufficiently large n it follows [4] by the perturbation theorem that

$$\|\hat{\underset{\sim}{\phi}} - \underset{\sim}{\phi}_0\| \le \|\underset{\sim}{M}^{-1}\underset{\sim}{E}\| \, \|\underset{\sim}{\phi}_0\|$$

while

$$\|\underset{\sim}{\phi}_0-\underset{\sim}{\phi}\| = \|\underset{\sim}{M}^{-1}\underset{\sim}{\varepsilon}\| \quad \text{and} \quad \|\underset{\sim}{\phi}_0\| \le \|\underset{\sim}{M}^{-1}\underset{\sim}{\varepsilon}\| + \|\underset{\sim}{\phi}\|$$

Thus

$$\|\hat{\underset{\sim}{\phi}} - \underset{\sim}{\phi}\| \le (\|\underset{\sim}{M}^{-1}\underset{\sim}{\varepsilon}\| + \|\phi\|)\|\underset{\sim}{M}^{-1}\|\|\underset{\sim}{E}\| + \|\underset{\sim}{M}^{-1}\|\|\underset{\sim}{\varepsilon}\|,$$

and observe that by the Frobenius norm estimation

$$E\|\underset{\sim}{E}\|^2 \le \sum_{i\neq j} E\varepsilon^2_{|i-j|} = O(1/n).$$

<u>Theorem 2:</u> Under the same conditions as in Theorem 1 the approximate asymptotic distribution of $\hat{\underset{\sim}{\phi}}$ is $N(\underset{\sim}{\phi}, \sigma^2/n\ \underset{\sim}{\Gamma}_p^{-1})$.

<u>Proof</u>: Observe that $E\tilde{\underset{\sim}{\lambda}} = \underset{\sim}{\lambda} + O(1/n)$. From [1] p. 489 and using a Taylor series expansion to one term

$$\lim_{n\to\infty} n\ \mathrm{Cov}(\tilde{\lambda}_k,\tilde{\lambda}_j) = \frac{\sum_{r=-\infty}^{\infty}[\rho_r\rho_{r+k-j}+\rho_{r-j}\rho_{r+k}+2\rho_j\rho_k\rho_r^2-2\rho_j\rho_r\rho_{r-k}-2\rho_k\rho_r\rho_{r-j}]}{\pi^2(1-\rho_k^2)^{1/2}(1-\rho_j^2)^{1/2}}, \tag{3.6}$$

from which it follows that $E(\tilde{\lambda}_k-\lambda_k)^2 \longrightarrow 0$, $n\to\infty$. Hence $\operatorname{plim}_{n\to\infty}\tilde{\underset{\sim}{\lambda}} = \underset{\sim}{\lambda}$ and by the fact that $\underset{\sim}{\phi}(\underset{\sim}{\lambda})$ is continuous also $\operatorname{plim}_{n\to\infty}\tilde{\underset{\sim}{\phi}} = \underset{\sim}{\phi}$. From this and Theorem 1 then

$$\operatorname{plim}_{n\to\infty}(\hat{\underset{\sim}{\phi}}-\tilde{\underset{\sim}{\phi}}) = \underset{\sim}{0}.$$

Therefore it is sufficient to consider the asymptotic distribution of $\tilde{\underset{\sim}{\phi}}$ only. Since it is well known that $\tilde{\underset{\sim}{\phi}}$ has an asymptotic normal distribution we only sketch the rest of the proof for convenience. The asymptotic covariance matrix can be obtained from

(3.1) by noting that the ij element of the information matrix is

$$-E \frac{\partial^2}{\partial\phi_i \partial\phi_j} \ell(\underset{\sim}{\phi}) = \frac{n-p}{\sigma^2} \gamma_{|i-j|}$$

and so $\mathrm{Var}(\tilde{\underset{\sim}{\phi}}) \approx \frac{\sigma^2}{n-p} \Gamma_p^{-1}$. Observe that

$$\sqrt{n}(\tilde{\underset{\sim}{\phi}} - \underset{\sim}{\phi}) = \sqrt{n}\left(\frac{1}{n-p} \sum_{t=p+1}^{n} \underset{\sim}{z}_t \underset{\sim}{z}_t'\right)^{-1} \frac{1}{n-p} \sum_{t=p+1}^{n} \underset{\sim}{z}_t u_t$$

and this has the same limiting distribution as $\underset{\sim}{y} = 1/\sqrt{n}\Gamma_p^{-1} \sum_{t=p+1}^{n} \underset{\sim}{z}_t u_t$. Anderson [1] pp. 198-200 has shown that if $\underset{\sim}{c}$ is any vector of constants then $\underset{\sim}{c}'\underset{\sim}{y}$ has an asymptotic normal distribution. It follows that $\sqrt{n}(\tilde{\underset{\sim}{\phi}} - \underset{\sim}{\phi}) \xrightarrow{L} N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{\Gamma}_p^{-1})$.

When γ_0, the variance of Z_t, is known then an approximate maximum likelihood estimator of σ^2 is,

$$\hat{\sigma}^2 = \gamma_0[1 - \sum_{k=1}^{p} \hat{\phi}_k \sin\pi(\hat{\lambda}_k - \tfrac{1}{2})] \tag{3.7}$$

However when γ_0 is unknown it is impossible to estimate it from the clipped data since the variability in the Z_t process is not preserved in the axis crossings. This means that there is no connection between λ_k and σ^2. In order to be able to estimate σ^2 the quantity Σz_t^2 is needed in addition to the clipped data.

We consider now two important special cases.

<u>Example 3.1. Estimation in AR(1).</u>

Assume $Z_t = \phi_1 Z_{t-1} + u_t$, $|\phi_1| < 1$. The likelihood (2.2) reduces to

$$L(\lambda_1) = 1/2(1 - \lambda_1)^d \lambda_1^{(n-1)-d}, \tag{3.8}$$

where $d = 2s - 2r_1 - (x_1 + x_n)$ is the number of axis crossings by Z_t. Owing to (2.9), (3.4) becomes

$$\hat{\phi}_1 = \phi_1(\hat{\lambda}_1) = \sin\pi\{\frac{(n-1) - (\#\text{ of axis crossings})}{n-1} - \frac{1}{2}\} \quad (3.9)$$

This estimator has been studied in [7] where it was shown that it competes well with $\tilde{\phi}_1$ the estimate obtained from the original data. $\hat{\phi}_1$ behaves remarkably well even for $|\phi_1|$ close to 1. Now from (3.8) and the uniqueness of the binomial expansion we immediately obtain (see [8] for another proof)

$$D \sim b(n-1,\ 1-\lambda_1). \quad (3.10)$$

That is, the distribution of the number of axis crossings D under the assumption that X_t is a first order Markov chain is binomial. However in [8] it was experimentally shown that (3.10) is reasonable only when ϕ_1 is in the approximate range $|\phi_1| \leq 0.6$. This discussion leads to the interesting observation that the Markov assumption is suitable for estimation as is seen from Theorem 1. However the distribution (3.10) obtained under this assumption should be regarded not without reservations. This seemingly a paradox can be resolved as follows. The estimation problem is rather simple since estimated is only one parameter while the distribution problem is much more complex and may involve many parameters which our Markov assumption excludes from consideration. But when $|\phi_1|$ is rather small (3.10) is adequate from which it follows directly that $\sqrt{n}(\hat{\lambda}_1 - \lambda_1) \xrightarrow{L} N(0, \lambda_1(1-\lambda_1))$, $n \to \infty$, which in turn implies that $\sqrt{n}(\hat{\phi}_1 - \phi_1) \xrightarrow{L} N(0, \pi^2\lambda_1(1-\lambda_1))$, $n \to \infty$. Thus, to test the hypothesis

that Z_t is white noise, i.e. $H_0:\phi_1 = 0$, we reject for large values of $\sqrt{n}\ |\hat{\phi}_1|/\frac{1}{2}\pi$.

The main advantage of $\hat{\phi}_1$ is that its computation is fast and can be carried on a small calculator. Let us compare the expected number of arithmetical operations needed in order to calculate $\hat{\phi}_1$ vs. the number needed for $\tilde{\phi}_1$. Neglecting end effects we have:

Expected # of arithmetical operations needed for $\hat{\phi}_1 \approx \frac{1}{2}(1-\lambda_1)n+6$

Expected # of arithmetical operations needed for $\tilde{\phi}_1 \approx 4n$,

where taking the sine and counting a 1-run are considered as single operations.

<u>Example 3.2: Estimation in AR(2)</u>

Here (1.2) becomes

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + u_t,$$

The estimates (3.4) are

$$\hat{\phi}_1 = \frac{\sin[\pi(\hat{\lambda}_1-\frac{1}{2})][1 - \sin\pi(\hat{\lambda}_2-\frac{1}{2})]}{\mathrm{Cos}^2[\pi(\hat{\lambda}_1-\frac{1}{2})]}$$

$$\hat{\phi}_2 = \frac{\sin\pi(\hat{\lambda}_2-\frac{1}{2}) - \sin^2[\pi(\hat{\lambda}_1-\frac{1}{2})]}{\mathrm{Cos}^2[\pi(\hat{\lambda}_1-\frac{1}{2})]} \qquad (3.11)$$

Note that $|\hat{\phi}_2| < 1$, $\hat{\phi}_1 + \hat{\phi}_2 < 1$, $\hat{\phi}_2 - \hat{\phi}_1 < 1$ as is needed for Z_t to be stationary.

It follows from (2.8), (2.10) that $\hat{\phi}_1$ and $\hat{\phi}_2$ are functions of the number of axis crossings and the number of runs with two or more symbols in the 0-1 series. We investigate here via an extensive simulation a case suggested by Wold (1965) and studied in [1] section 5.9. Accordingly $\phi_1 = \phi$, $\phi_2 = -\phi^2$, $\sigma^2 = (1 - \phi^6)/(1 + \phi^2)$ so that $\mathrm{Var}Z_t = 1$. The model is then

$$Z_t = \phi Z_{t-1} - \phi^2 Z_{t-2} + u_t, \quad \mathrm{Var}Z_t = 1.$$

The values of ϕ are 0.25, 0.7, 0.9 and of n 250, 500, 1000, 2000. For each pair (ϕ,n) ten time series were generated from which the sample averages and variances of $\hat{\phi}_1$, $\hat{\phi}_2$ and of $\tilde{\phi}_1$, $\tilde{\phi}_2$ were computed. The results are given in Table 2. Table 1 gives typical records corresponding to the three choices of ϕ. It is seen that $\hat{\phi}_1$, $\hat{\phi}_2$ are on the average just as good point estimates as $\tilde{\phi}_1$, $\tilde{\phi}_2$ are. However the variances of $\tilde{\phi}_1$, $\tilde{\phi}_2$ are usually smaller than those of $\hat{\phi}_1$ and $\hat{\phi}_2$. Roughly speaking, in order to achieve the same order of precision for the estimates based on the clipped data, the 0-1 series should be approximately twice as large as the original series. Although the estimates based on the original series are endowed with a relatively higher efficiency, for large data records the difference between these two types of estimators is negligible. This is in accordance with Theorem 2. The study of this relative efficiency is the subject of the next section.

What happens if in the clipped series the 0's are replaced by 1's and the 1's are replaced by 0's? In this case the information pertaining to axis crossings is left intact and the

same is true for the number of runs with at least two symbols between the first and last 1 except for negligible end effects. This means that we obtain almost exactly the same estimates for ϕ_1, ϕ_2. For example with $n = 1000$ and $\phi = 0.25$ we obtained ave $\hat{\phi}_1 = 0.2539$, ave $\hat{\phi}_2 = -0.0646$, Var $\hat{\phi}_1 = 0.0017$, Var $\hat{\phi}_2 = 0.0026$ which are almost the same as the corresponding results in Table 2.

Table 1. Typical realizations of size 50 obtained by clipping

$$Z_t = \phi Z_{t-1} - \phi^2 Z_{t-2} + u_t,$$

$$u_t \sim N(0,(1-\phi^6)/(1+\phi^2))$$

ϕ	Clipped Series	# Axis-crossings
0.25	01111010100001101111001110110000100110001011100011	23
0.7	01111000110001100111001110010000110010001111100011	19
0.9	11000111000111000111000111000100000000111000111110	15

Again (3.11) can be computed much faster than $\tilde{\phi}_1$ and $\tilde{\phi}_2$. In fact the difference in the expected number of operationsis approximately $6n - \frac{1}{2}(2 - \lambda_1 - \lambda_2)$ in favor of $\hat{\phi}_1$ and $\hat{\phi}_2$.

Table 2. Estimates of the Coefficients in AR(2).

ϕ_1	ϕ_2	n	ave $\hat{\phi}_1$	ave $\hat{\phi}_2$	ave $\tilde{\phi}_1$	ave $\tilde{\phi}_2$	Var $\hat{\phi}_1$	Var $\hat{\phi}_2$	Var $\tilde{\phi}_1$	Var $\tilde{\phi}_2$
0.25	-0.0625	250	0.2263	-0.1011	0.2280	-0.0626	0.0089	0.0059	0.0048	0.0043
		500	0.2454	-0.0621	0.2416	-0.0666	0.0048	0.0032	0.0037	0.0017
		1000	0.2520	-0.0625	0.2541	-0.0615	0.0016	0.0026	0.0013	0.0019
		2000	0.2593	-0.0691	0.2583	-0.0733	0.0017	0.0013	0.0008	0.0009
0.7	-0.49	250	0.6878	-0.4944	0.6867	-0.4945	0.0184	0.0071	0.0016	0.0015
		500	0.6922	-0.4857	0.6980	-0.4926	0.0034	0.0026	0.0006	0.0004
		1000	0.6955	-0.4833	0.7025	-0.4902	0.0013	0.0023	0.0005	0.0007
		2000	0.7028	-0.4680	0.7000	-0.4836	0.0002	0.0011	0.0003	0.0003
0.9	-0.81	250	0.9222	-0.7901	0.8905	-0.8037	0.0054	0.0043	0.0020	0.0012
		500	0.9313	-0.8164	0.9094	-0.8165	0.0037	0.0019	0.0003	0.0005
		1000	0.9083	-0.8044	0.9061	-0.8064	0.0018	0.0012	0.0002	0.0004
		2000	0.9003	-0.8058	0.9012	-0.8080	0.0009	0.0009	0.0001	0.0002

§4. Loss of efficiency

A problem associated with estimation after hard limiting is the problem of measuring the loss in the precision of our estimates due to this transformation. That clipping results in a loss of "information" can be seen from the inequality

$$|\rho_x(k)| \le |\rho_k| \tag{4.1}$$

where $\rho_x(k) = \frac{2}{\pi}\sin^{-1}(\rho_k)$ is the autocorrelation of X_t. Now since in addition X_t and X_{t-k} are independent if and only if they are uncorrelated, (4.1) means that the pairwise dependence in the clipped series is weaker than that in Z_t. Intuitively, this may affect the efficiency of estimates computed from the clipped data. More precisely, if $\phi_\upsilon(\underset{\sim}{\lambda})$ is the υth component of $\underset{\sim}{\phi}(\underset{\sim}{\lambda})$ let $a_{i\upsilon} = \partial\phi_\upsilon(\underset{\sim}{\lambda})/\partial\lambda_i$. Then upon using a Taylor series approximation we obtain when n is sufficiently large

$$\begin{aligned} \operatorname{Var}\hat{\phi}_\upsilon &= \operatorname{Var}\phi_\upsilon(\hat{\underset{\sim}{\lambda}}) \approx \underset{\sim}{a}_\upsilon' \underset{\sim}{\Sigma}_{\hat{\lambda}} \underset{\sim}{a}_\upsilon \\ \operatorname{Var}\tilde{\phi}_\upsilon &= \operatorname{Var}\phi_\upsilon(\tilde{\underset{\sim}{\lambda}}) \approx \underset{\sim}{a}_\upsilon' \underset{\sim}{\Sigma}_{\tilde{\lambda}} \underset{\sim}{a}_\upsilon \end{aligned} \tag{4.2}$$

where $\underset{\sim}{a}_\upsilon = (a_{1\upsilon},\ldots,a_{p\upsilon})$ and $\underset{\sim}{\Sigma}_{\hat{\lambda}}$ is the covariance matrix of $\hat{\underset{\sim}{\lambda}}$. It is convenient to associate the efficiency of $\hat{\underset{\sim}{\phi}}$ with the sum of the eigenvalues of $\underset{\sim}{\Sigma}_{\hat{\lambda}}$. Accordingly we define

$$\operatorname{eff}(\hat{\underset{\sim}{\phi}}) \overset{\text{def}}{=\!=} \frac{\operatorname{tr}\underset{\sim}{\Sigma}_{\hat{\lambda}}}{\operatorname{tr}\underset{\sim}{\Sigma}_{\tilde{\lambda}}} \tag{4.3}$$

The choice for (4.3) stems from the simplicity of its calculation and the fact that if $\mathrm{tr}\underset{\sim}{\Sigma}_{\hat{\lambda}}$ is relatively large so will be $\mathrm{Var}\ \hat{\phi}_\upsilon$ too. Using the dominating terms in (3.5) and (3.6) we obtain

$$\frac{\mathrm{Var}\ \hat{\lambda}_k}{\mathrm{Var}\ \tilde{\lambda}_k} \approx \frac{(1-\rho_k^2)\ \sum_{r=-\infty}^{\infty}(\sin^{-1}(\rho_r))^2}{\sum_{r=-\infty}^{\infty}\rho_r^2} \tag{4.4}$$

Therefore when ρ_k is <u>small</u>, e.g. $|\rho_k| \leq 0.2$, $k \neq 0$, (4.3) is near unity and the clipping results in almost no loss for all practical purposes. In general we obtain an approximate upper bound using (4.1)

$$\mathrm{eff}(\hat{\underset{\sim}{\phi}}) \leq \max_{k=1,\ldots,p}\ (1-\rho_k^2)\ \frac{\pi^2}{4} \tag{4.5}$$

Because $\mathrm{Var}\ \hat{\phi}_\upsilon = O(1/n)$ (4.5) means that for $\hat{\phi}_\upsilon$ to be as efficient as $\tilde{\phi}_\upsilon$ is the binary series should be roughly twice as long as the original series. The experimental results in Example (3.2) support this finding. See [13], [18] for similar results. It should be noted however that for long series the difference between $\hat{\underset{\sim}{\phi}}$ and $\tilde{\underset{\sim}{\phi}}$ is negligible. Also, observe that when $\rho_k \geq 0$ an approximate upper bound for the ratio $\mathrm{Var}\ \hat{\phi}_\upsilon / \mathrm{Var}\ \tilde{\phi}_\upsilon$ is again (4.5). Finally it is easy to see that the expected <u>number of operations that can be saved owing to clipping is at least $(p+2)n$ for an AR(p).</u>

References

[1] Anderson, T.W., The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc., 1971.

[2] Bartlett, M.S., "The Frequency Goodness of Fit Test for Probability Chains," Proc. Camb. Phil. Soc., 47, (1951), 86-95.

[3] Box, G.E.P. and Jenkins, G.M., Time Series Analysis Forecasting and Control, San Francisco: Holden-Day, 1970.

[4] Dancis, J. Lecture Notes in Linear Algebra, University of Maryland, 1977.

[5] Durbin, J., "The Fitting of Time Series Models," Rev. Int. Inst. Stat., 28, (June 1960), 233-244.

[6] Hinich, M., "Estimation of Spectra After Hard Clipping of Gaussian Processes," Technometrics, 9, (August 1967), 391-400.

[7] Kedem, B., "Exact Maximum Likelihood Estimation of the AR(1) Process After Hard Limiting," IEEE Trans. Inf. Theory, IT-22, (July 1976), 491-493.

[8] Kedem, B., "Sufficiency and the Number of Level Crossings by a Stationary Process," To appear.

[9] Lawson, J.L. and Uhlenbeck, G.E., Threshold Signals, New York: Dover Publ. Inc., 1950.

[10] Lindgren, G., "Spectral Moments Estimation by Means of Level Crossings," Biometrika, 61, (December 1974), 401-418.

[11] Lomnicki, Z.A. and Zaremba, S.K., "Some Applications of Zero-One Processes," Jour. Roy. Stat. Soc. B, 17, No. 2 (1955), 243-255.

[12] Mann, H.B. and Wald, A., "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrika, 11, (july 1943), 173-220.

[13] McNeil, D.R., "Estimating the Covariance and Spectral Density Functions From Clipped Stationary Time Series," Jour. Roy. Stat. Soc. B, 29, No. 1 (1967), 180-195.

[14] Parzen, E., "An Approach to Time Series Modeling and Forecasting Illustrated by Hourly Electricity Demands," Technical Report No. 37, SUNY at Buffalo, (May 1976).

[15] Sheppard, W.F., "On the Application of the Theory of Error to the Cases of Normal Distribution and Normal Correlation," Philosophical Trans. Roy. Soc. of London, A, 192, (1899), 101-167.

[16] Stieltjes, T.J., "Extrait D'une Lettre Adressée à M. Hermite," Bull. Sci. Math., 2:E Série, 13, (1889), 170-172.

[17] Van Vleck, J., "The Spectrum of Clipped Noise," Radio Research Lab. Report No. 51, Harvard Univ., (July 1943).

[18] Weinreb, S., "A Digital Spectral Analysis Technique and its Application to Radio Astronomy," Technical Report No. 412, M.I.T. Res. Lab. of Electronics, (May 1963).

This research was supported by the Air Force Office of Scientific Research Grant AFOSR 77-3332.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 77-1231	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATION OF THE PARAMETERS IN STATIONARY AUTOREGRESSIVE PROCESSES AFTER HARD LIMITING		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER TR 77-42
7. AUTHOR(s) Benjamin Kedem		8. CONTRACT OR GRANT NUMBER(s) AFOSR 77-3332 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Maryland Department of Mathematics College Park, MD 20742		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332		12. REPORT DATE August 1977
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

binary, axis-crossings, maximum likelihood, conditional probabilities, perturbation, efficiency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The parameters of a stationary AR(p) process are estimated after clipping. This estimation is based in part on the number of certain runs in the binary series. Very little precision is lost due to this quantization but the expected number of arithmetical operations which are saved is at least (p+2)n where counting a run is considered as an operation and n is the series size.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)